```
import pandas as pd
```

```
df = pd.read_csv("spotify_millsongdata.csv")
```

```
df.head(5)
```

	artist	song	link	text
0	ABBA	Ahe's My Kind Of Girl	/a/abba/ahes+my+kind+of+girl_20598417.html	Look at her face, it's a wonderful face \r\nA...
1	ABBA	Andante, Andante	/a/abba/andante+andante_20002708.html	Take it easy with me, please \r\nTouch me gen...
2	ABBA	As Good As New	/a/abba/as+good+as+new_20003033.html	I'll never know why I had to go \r\nWhy I had...
3	ABBA	Bang	/a/abba/bang_20598415.html	Making somebody happy is a question of give an...
4	ABBA	Bang-A-Boomerang	/a/abba/bang+a+boomerang_20002668.html	Making somebody happy is a question of give an...

```
df.tail(5)
```

	artist	song	link	text
57645	Ziggy Marley	Good Old Days	/z/ziggy+marley/good+old+days_10198588.html	Irie days come on play \r\nLet the angels fly...
57646	Ziggy Marley	Hand To Mouth	/z/ziggy+marley/hand+to+mouth_20531167.html	Power to the workers \r\nMore power \r\nPowe...
57647	Zwan	Come With Me	/z/zwan/come+with+me_20148981.html	all you need \r\nis something i'll believe \...
57648	Zwan	Desire	/z/zwan/desire_20148986.html	northern star \r\nam i frightened \r\nwhere ...
57649	Zwan	Heartsong	/z/zwan/heartsong_20148991.html	come in \r\nmake yourself at home \r\ni'm a ...

```
df.shape
```

```
(57650, 4)
```

```
df.isnull().sum()
```

```
artist    0
song      0
link      0
text      0
dtype: int64
```

```
df =df.sample(5000).drop('link', axis=1).reset_index(drop=True)
```

```
df.head(10)
```

	artist	song	text
0	ABBA	Ahe's My Kind Of Girl	Look at her face, it's a wonderful face \r\nA...
1	ABBA	Andante, Andante	Take it easy with me, please \r\nTouch me gen...
2	ABBA	As Good As New	I'll never know why I had to go \r\nWhy I had...
3	ABBA	Bang	Making somebody happy is a question of give an...
4	ABBA	Bang-A-Boomerang	Making somebody happy is a question of give an...
5	ABBA	Burning My Bridges	Well, you hoot and you holler and you make me ...
6	ABBA	Cassandra	Down in the street they're all singing and sho...
7	ABBA	Chiquitita	Chiquitita, tell me what's wrong \r\nYou're e...
8	ABBA	Crazy World	I was out with the morning sun \r\nCouldn't s...
9	ABBA	Crying Over You	I'm waitin' for you baby \r\nI'm sitting all ...

```
df['text'][0]
```

```
"Look at her face, it's a wonderful face  \r\nAnd it means something special to me  \r\nLook at the way that she smiles when she sees me  \r\nHow lucky can one fellow be?  \r\n  \r\nShe's just my kind of girl, she makes me feel fine  \r\nWho could ever believe that she could be mine?  \r\nShe's just my kind of girl, without her I'm blue  \r\nAnd if she ever leaves me what could I do, what could I do?  \r\n  \r\nAnd when we go for a walk in the park  \r\nAnd she holds me and squeezes my hand  \r\nWe'll go on walking for hours and
```

```
talking  \r\nAbout all the things that we plan  \r\n  \r\nShe's just my kind of girl, she makes me feel fine  \r\nWho could ever believe that she could be mine?  \r\nShe's just my kind of girl, without her I'm blue  \r\nAnd if she ever leaves me what could I do, what could I do?\r\n\r\n"
```

```
# df = df.sample(5000)
```

```
df.shape
```

```
(5000, 3)
```

Text Cleaning/ Text Preprocessing

```
df['text'] = df['text'].str.lower().replace(r'^\w\s', ' ').replace(r'\n', ' ', regex = True)
```

```
import nltk
from nltk.stem.porter import PorterStemmer
stemmer = PorterStemmer()

def tokenization(txt):
    tokens = nltk.word_tokenize(txt)
    stemming = [stemmer.stem(w) for w in tokens]
    return " ".join(stemming)
```

```
df['text'] = df['text'].apply(lambda x: tokenization(x))
```

```
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.metrics.pairwise import cosine_similarity
```

```
tfidvector = TfidfVectorizer(analyzer='word',stop_words='english')
matrix = tfidvector.fit_transform(df['text'])
similarity = cosine_similarity(matrix)
```

```
similarity[0]
```

```
df[df['song'] == 'Crying Over You']
```

```
def recommendation(song_df):
    idx = df[df['song'] == song_df].index[0]
    distances = sorted(list(enumerate(similarity[idx])),reverse=True,key=lambda x:x[1])

    songs = []
    for m_id in distances[1:21]:
        songs.append(df.iloc[m_id[0]].song)

    return songs
```

```
recommendation('Crying Over You')
```

```
import pickle
pickle.dump(similarity,open('similarity.pkl','wb'))
pickle.dump(df,open('df.pkl','wb'))
```